

Supplementary Material of “Robust Multi-view Subspace Learning with Non-identically and Non-independently Distributed Complex Noises”

Zongsheng Yue, Hongwei Yong, Deyu Meng, *Member, IEEE*, Qian Zhao, Yee Leung,
and Lei Zhang, *Fellow, IEEE*

Abstract—In this supplementary material, we provide more details on the computations involved in the proposed variational inference algorithm and more experiment results.

I. NIID-MSL MODEL

A. Model Formulation

Basically, we decompose the observed data into:

$$\mathbf{X}^v = \mathbf{S}^v + \mathbf{E}^v, \quad (1)$$

where $\mathbf{E}^v = \{e_{ij}^v\}_{d \times n}$ denotes the residual term (i.e., noise component) and $\mathbf{S}^v \in \mathbb{R}^{d \times n}$ is the expected data located on the latent subspace, d and n represent the dimensionality and the number of samples in each view, respectively.

Firstly, we model the noise term \mathbf{E}^v as follows:

$$\xi_k \sim \text{Gam}(e_0, f_0), \quad e_{ij}^v \sim \mathcal{N}(0, (\xi_{c_{z_{ij}}^v})^{-1}), \quad (2a)$$

$$c_t^v \sim \text{Multi}(\boldsymbol{\beta}), \quad z_{ij}^v \sim \text{Multi}(\boldsymbol{\pi}^v), \quad (2b)$$

$$\beta_k = \beta'_k \prod_{l=1}^{k-1} (1 - \beta'_l), \quad \pi_t^v = \pi_t^{v'} \prod_{s=1}^{t-1} (1 - \pi_s^{v'}), \quad (2c)$$

$$\beta'_k \sim \text{Beta}(1, \gamma), \quad \pi_t^{v'} \sim \text{Beta}(1, \alpha^v), \quad (2d)$$

$$\gamma \sim \text{Gam}(m_0, n_0), \quad \alpha^v \sim \text{Gam}(g_0, h_0). \quad (2e)$$

where $\boldsymbol{\alpha}$ and γ are the concentration parameters, which mainly affect the number of Gaussian components of the second-level GMM in each view and the first-level GMM for the entire dataset, respectively.

As for the expected data \mathbf{S}^v , we embed each view into a latent space \mathbf{R} with a dictionary \mathbf{L}^v as in the conventional MSL methods, i.e.,

$$\mathbf{S}^v = \sum_{r=1}^l \mathbf{L}_{r \cdot}^v \mathbf{R}_{r \cdot}, \quad (3a)$$

$$\mathbf{R}_{r \cdot} \sim \mathcal{N}(\mathbf{0}, \frac{1}{\tau_r} I_n), \quad (3b)$$

$$\mathbf{L}_{r \cdot}^v \sim \mathcal{N}(\mathbf{0}, \frac{1}{\lambda_r^v} I_d), \quad (3c)$$

$$\lambda_r^v \sim \text{Gam}(a_0, b_0), \quad \tau_r \sim \text{Gam}(c_0, d_0). \quad (3d)$$

Combining Eqs. (1) - (3), the goal of our proposed NIID-MSL turns to infer the posteriors of all involved variables:

$$p(\mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \gamma | \mathcal{X}), \quad (4)$$

where $\mathbf{C} = \{c_t^v\}$, $\mathbf{Z} = \{z_{ij}^v\}$.

B. Variational Assumption

The full likelihood of the proposed NIID-MSL model is expressed as:

$$\begin{aligned} & p(\mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \gamma, \mathbf{X}) \\ &= p(\mathbf{X} | \mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}) p(\mathbf{L} | \boldsymbol{\lambda}) p(\boldsymbol{\lambda}) p(\mathbf{R} | \boldsymbol{\tau}) p(\boldsymbol{\tau}) p(\boldsymbol{\xi}) p(\mathbf{C} | \boldsymbol{\beta}) \\ &\quad p(\boldsymbol{\beta}' | \gamma) p(\gamma) p(\mathbf{Z} | \boldsymbol{\pi}) p(\boldsymbol{\pi} | \boldsymbol{\alpha}) p(\boldsymbol{\alpha}) \\ &= \prod_{v,i,j} \prod_t \left\{ \prod_k \mathcal{N}\left(x_{ij}^v | \mathbf{L}_{i \cdot}^v \mathbf{R}_{j \cdot}, \xi_k^{-1}\right)^{\mathbf{1}[c_t^v=k]} \right\}^{\mathbf{1}[z_{ij}^v=t]} \\ &\quad \prod_{v,i,j} \text{Multi}(z_{ij}^v | \boldsymbol{\pi}^v) \prod_{v,t} \text{Multi}(c_t^v | \boldsymbol{\beta}) \prod_k \text{Gam}(\xi_k | e_0, f_0) \\ &\quad \prod_{v,r} \mathcal{N}\left(\mathbf{L}_r^v | \mathbf{0}, \lambda_r^{v-1} I_m\right) \text{Gam}(\lambda_r^v | a_0, b_0) \\ &\quad \prod_r \mathcal{N}(\mathbf{R}_{r \cdot} | \mathbf{0}, \tau_r^{-1} I_n) \text{Gam}(\tau_r | c_0, d_0) \\ &\quad \prod_{v,t} \text{Beta}(\pi_t^{v'} | 1, \alpha^v) \prod_v \text{Gam}(\alpha^v | m_0, n_0) \\ &\quad \prod_k \text{Beta}(\beta'_k | 1, \gamma) \text{Gam}(\gamma | g_0, h_0). \end{aligned} \quad (5)$$

In the main text, we have introduced the variational inference to calculate the posterior of this model and assumed that the approximation of the posterior has a factorized form as follows:

$$\begin{aligned} & q(\mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \gamma) \\ &= \prod_{i=1}^d q(\mathbf{L}_{i \cdot}^v | \boldsymbol{\mu}_i^v, \boldsymbol{\Sigma}_i^v) \prod_{j=1}^n q(\mathbf{R}_{j \cdot} | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \prod_{k=1}^K q(\xi_k | e_k, f_k) \\ &\quad \prod_{v=1}^V \prod_{i,j}^d q(z_{ij}^v | \boldsymbol{\rho}_{ij}^v) \prod_{v=1}^V q(\alpha^v | m^v, n^v) q(\gamma | g, h) \\ &\quad \prod_{v=1}^V \prod_{t=1}^T q(c_t^v | \boldsymbol{\varphi}_t^v) q(\pi_t^{v'} | r_t^v, w_t^v) \prod_{k=1}^K q(\beta'_k | s_k^1, s_k^2) \\ &\quad \prod_{v=1}^V \prod_{r=1}^l q(\lambda_r^v | a_r^v, b_r^v) \prod_{r=1}^l q(\tau_r | c_r, d_r). \end{aligned} \quad (6)$$

Next, we give the detailed deduction of each factorized distribution involved in the posterior of Eq. (6). $E_{\mathbf{x} \setminus x_i}[f(\mathbf{x})]$ denotes the expectation of $f(\mathbf{x})$ on the set of \mathbf{x} with x_i

removed. For notation convenience, we introduce Θ to denote all the parameters that need to be inferred, i.e.,

$$\Theta = \{\mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \boldsymbol{\gamma}\}.$$

Infer C and Z :

$$\begin{aligned} & \ln q(z_{ij}^v) \\ &= E_{\Theta \setminus Z} [p(\mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \mathbf{X})] + const \\ &= \sum_t \mathbf{1}[z_{ij}^v = t] \left\{ \sum_k \varphi_{tk}^v E_{\Theta \setminus Z} [\mathcal{N}(x_{ij}^v - \mathbf{L}_{i \cdot}^v \mathbf{R}_{\cdot j} | 0, \xi_k^{-1})] \right. \\ &\quad \left. + E[\ln \pi_t^v] \right\} + const \\ &= \sum_t \mathbf{1}[z_{ij}^v = t] \left\{ \sum_k \varphi_{tk}^v \left(-\frac{1}{2} \ln 2\pi + \frac{1}{2} E[\ln \xi_k] \right. \right. \\ &\quad \left. \left. - \frac{1}{2} E[\xi_k] E[(x_{ij}^v - \mathbf{L}_{i \cdot}^v \mathbf{R}_{\cdot j})^2] \right) + E[\ln \pi_t^v] \right\} + const, \end{aligned} \quad (7)$$

$$\begin{aligned} & \ln q(c_t^v) \\ &= E_{\Theta \setminus C} [p(\mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \mathbf{X})] + const \\ &= \sum_k \mathbf{1}[c_t^v = k] \left\{ \sum_{i,j} \rho_{ijt}^v E_{\Theta \setminus C} [\mathcal{N}(x_{ij}^v - \mathbf{L}_{i \cdot}^v \mathbf{R}_{\cdot j} | 0, \xi_k^{-1})] \right. \\ &\quad \left. + E[\ln \beta_k] \right\} + const \\ &= \sum_k \mathbf{1}[c_t^v = k] \left\{ \sum_{i,j} \rho_{ijt}^v \left(-\frac{1}{2} \ln 2\pi + \frac{1}{2} E[\ln \xi_k] \right. \right. \\ &\quad \left. \left. - \frac{1}{2} E[\xi_k] E[(x_{ij}^v - \mathbf{L}_{i \cdot}^v \mathbf{R}_{\cdot j})^2] \right) + E[\ln \beta_k] \right\} + const, \end{aligned} \quad (8)$$

Taking the exponential of both sides of Eq. (7), Eq. (8) and normalizing the right hand side, we obtain

$$q(z_{ij}^v | \rho_{ij}^v) = \text{Multi}(\rho_{ij}^v), \quad q(c_t^v | \varphi_t^v) = \text{Multi}(\varphi_t^v), \quad (9)$$

where

$$\rho_{ijt}^v = \frac{\rho_{ijt}^{v'}}{\sum_s \rho_{ijs}^{v'}}, \quad \varphi_{tk}^v = \frac{\varphi_{tk}^{v'}}{\sum_s \varphi_{ts}^{v'}}, \quad (10)$$

$$\begin{aligned} \rho_{ijt}^{v'} &\propto \exp \left\{ \sum_k \varphi_{tk}^{v'} \left(\frac{1}{2} \ln 2\pi + \frac{1}{2} E[\ln \xi_k] \right. \right. \\ &\quad \left. \left. - \frac{1}{2} E[\xi_k] E[(x_{ij}^v - \mathbf{L}_{i \cdot}^v \mathbf{R}_{\cdot j})^2] \right) + E[\ln \pi_t^v] \right\}, \end{aligned} \quad (11)$$

$$\begin{aligned} \varphi_{tk}^{v'} &\propto \exp \left\{ \sum_{i,j} \rho_{ijt}^v \left(\frac{1}{2} \ln 2\pi + \frac{1}{2} E[\ln \xi_k] \right. \right. \\ &\quad \left. \left. - \frac{1}{2} E[\xi_k] E[(x_{ij}^v - \mathbf{L}_{i \cdot}^v \mathbf{R}_{\cdot j})^2] \right) + E[\ln \beta_k] \right\}. \end{aligned} \quad (12)$$

Infer ξ :

$$\begin{aligned} & \ln q(\xi_k) \\ &= E_{\Theta \setminus \xi} [p(\mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \mathbf{X})] + const \\ &= \sum_{v,i,j,t} \rho_{ijt}^v \varphi_{tk}^v E_{\Theta \setminus \xi_k} [\mathcal{N}(x_{ij}^v - \mathbf{L}_{i \cdot}^v \mathbf{R}_{\cdot j} | 0, \xi_k^{-1})] \\ &\quad + (e_0 - 1) \ln \xi_k - f_0 \xi_k \\ &= \left(\frac{1}{2} \sum_{v,i,j,t} \rho_{ijt}^v \varphi_{tk}^v + e_0 - 1 \right) \ln \xi_k \\ &\quad - \left\{ \frac{1}{2} \sum_{v,i,j,t} \rho_{ijt}^v \varphi_{tk}^v E[(x_{ij}^v - \mathbf{L}_{i \cdot}^v \mathbf{R}_{\cdot j})^2] + f_0 \right\} \xi_k + const. \end{aligned} \quad (13)$$

After taking the exponential of both sides of Eq. (13), we have:

$$q(\xi_k | e_k, f_k) = \text{Gam}(\xi_k | e_k, f_k), \quad (14)$$

where

$$e_k = \frac{1}{2} \sum_{v,i,j,t} \rho_{ijt}^v \varphi_{tk}^v + e_0, \quad (15)$$

$$f_k = \frac{1}{2} \sum_{v,i,j,t} \rho_{ijt}^v \varphi_{tk}^v E[(x_{ij}^v - \mathbf{L}_{i \cdot}^v \mathbf{R}_{\cdot j})^2] + f_0. \quad (16)$$

Infer π' and β' :

$$\begin{aligned} & \ln q(\pi_t^{v'}) \\ &= E_{\Theta \setminus \pi'} [p(\mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \mathbf{X})] + const \\ &= \sum_{i,j} \rho_{ijt}^v \ln \pi_t^v + (E[\alpha^v] - 1) \ln(1 - \pi_t^{v'}) + const \\ &= \left(\sum_{i,j,s=t+1} \rho_{ijs}^v + E[\alpha^v] - 1 \right) \ln(1 - \pi_t^{v'}) \\ &\quad + \left(\sum_{i,j} \rho_{ijt}^v \right) \ln \pi_t^{v'} + const. \end{aligned} \quad (17)$$

Then we take the exponential of both sides of Eq. (17) and get:

$$q(\pi_r^{v'} | r_t^v, w_t^v) = \text{Beta}(\pi_r^{v'} | r_t^v, w_t^v), \quad (18)$$

where

$$r_t^v = \sum_{i,j} \rho_{ijt}^v + 1, \quad (19)$$

$$w_t^v = \sum_{i,j,s=t+1} \rho_{ijs}^v + E[\alpha^v]. \quad (20)$$

Similarly, we have:

$$q(\beta_k' | s_k^1, s_k^2) = \text{Beta}(\beta_k' | s_k^1, s_k^2), \quad (21)$$

where

$$s_k^1 = \sum_{v,t} \varphi_{tk}^v + 1, \quad (22)$$

$$s_k^2 = \sum_{v,t,l=k+1} \varphi_{tl}^v + E[\gamma]. \quad (23)$$

Infer α and γ :

$$\begin{aligned} & \ln q(\alpha^v) \\ &= E_{\Theta \setminus \alpha} [p(\mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \mathbf{X})] + const \\ &= \sum_t \left((\alpha^v - 1) E[\ln(1 - \pi_t^{v'})] + \ln \alpha^v \right) + (m_0 - 1) \ln \alpha^v \\ &\quad - n_0 \alpha^v + const \\ &= (T + m_0 - 1) \ln \alpha^v - \left(n_0 - \sum_t E[\ln(1 - \pi_t^{v'})] \right) + const. \end{aligned} \quad (24)$$

From Eq. (24), We can easily get the following equations of α :

$$q(\alpha^v | m^v, n^v) = \text{Gam}(\alpha^v | m^v, n^v), \quad (25)$$

where

$$m^v = T + m_0, \quad (26)$$

$$n^v = n_0 - \sum_t E[\ln(1 - \pi_t^{v'})]. \quad (27)$$

Similarly, we can update the variable γ as follows:

$$q(\gamma | g, h) = \text{Gam}(\gamma | g, h), \quad (28)$$

where

$$g = K + g_0, \quad (29)$$

$$h = h_0 - \sum_k E[\ln(1 - \beta_k')]. \quad (30)$$

Infer L and R :

$$\begin{aligned} & \ln q(\mathbf{L}_i^v) \\ &= E_{\Theta \setminus L} [p(\mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \mathbf{X})] + const \\ &= \sum_{j,t,k} \rho_{ijt}^v \varphi_{tk}^v \left[-\frac{1}{2} E[\xi_k] E \left[\left(x_{ij}^v - \mathbf{L}_{i \cdot}^v \mathbf{R}_{j \cdot}^v \right)^2 \right] \right] \\ &\quad - \frac{1}{2} \mathbf{L}_{i \cdot}^v \mathbf{A}_v^L \mathbf{L}_{i \cdot}^v + const, \end{aligned} \quad (31)$$

where $\mathbf{A}_v^L = \text{diag}(E[\boldsymbol{\lambda}^v])$. Taking the exponential of both sides of Eq. (31), and normalizing the result, we obtain the posterior distribution of \mathbf{L}_i^v :

$$q(\mathbf{L}_{i \cdot}^v | \boldsymbol{\mu}_i^v, \boldsymbol{\Sigma}_i^v) = \mathcal{N}(\mathbf{L}_{i \cdot}^v | \boldsymbol{\mu}_i^v, \boldsymbol{\Sigma}_i^v), \quad (32)$$

where

$$\boldsymbol{\Sigma}_i^v = \left(\sum_{j,t,k} \rho_{ijt}^v \varphi_{tk}^v E[\xi_k] E[\mathbf{R}_{j \cdot} \mathbf{R}_{j \cdot}^T] + \mathbf{A}_v^L \right)^{-1}, \quad (33)$$

$$\boldsymbol{\mu}_i^v = \boldsymbol{\Sigma}_i^v \sum_{j,t,k} \rho_{ijt}^v \varphi_{tk}^v E[\xi_k] x_{ij}^v E[\mathbf{R}_{j \cdot}]. \quad (34)$$

Similarly, each column of \mathbf{R} is also a Gaussian distribution, i.e.,

$$q(\mathbf{R}_{\cdot j} | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) = \mathcal{N}(\mathbf{R}_{\cdot j} | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j), \quad (35)$$

where

$$\boldsymbol{\Sigma}_j = \left(\sum_{v,i,t,k} \rho_{ijt}^v \varphi_{tk}^v E[\xi_k] E[\mathbf{L}_{i \cdot}^v \mathbf{L}_{i \cdot}^{v \cdot T}] + \mathbf{A}^R \right)^{-1}, \quad (36)$$

$$\boldsymbol{\mu}_j = \boldsymbol{\Sigma}_j \sum_{v,i,t,k} \rho_{ijt}^v \varphi_{tk}^v E[\xi_k] x_{ij}^v E[\mathbf{L}_{i \cdot}^v]. \quad (37)$$

and $\mathbf{A}^R = \text{diag}(E[\boldsymbol{\tau}])$.

Infer λ and τ :

$$\begin{aligned} & \ln q(\lambda_r^v) \\ &= E_{\Theta \setminus \lambda} [p(\mathbf{L}, \mathbf{R}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \mathbf{X})] + const \\ &= \left(\frac{d}{2} + a_0 - 1 \right) \ln \lambda_r^v - \left(\frac{1}{2} E \left[\mathbf{L}_{r \cdot}^v \mathbf{L}_{r \cdot}^{v \cdot T} \right] + b_0 \right) \lambda_r^v + const. \end{aligned} \quad (38)$$

Thus, we can get the following updating equations:

$$q(\lambda_r^v | a_r^v, b_r^v) = \text{Gam}(\lambda_r^v | a_r^v, b_r^v), \quad (39)$$

where

$$a_r^v = \frac{d}{2} + a_0, \quad (40)$$

$$b_r^v = \frac{1}{2} E \left[\mathbf{L}_{r \cdot}^v \mathbf{L}_{r \cdot}^{v \cdot T} \right] + b_0. \quad (41)$$

Similarly, we can update τ as follows:

$$q(\tau_r | c_r, d_r) = \text{Gam}(\tau_r | c_r, d_r), \quad (42)$$

where

$$c_r = \frac{n}{2} + c_0, \quad (43)$$

$$d_r = \frac{1}{2} E \left[\mathbf{R}_{r \cdot}^T \mathbf{R}_{r \cdot} \right] + d_0. \quad (44)$$

C. Calculation of Expectations

The expectations in the variational update equations can be calculated with respect to the current variational distributions, as listed in the following:

$$E[\xi_k] = \frac{e_k}{f_k}, \quad (45)$$

$$E[\ln \xi_k] = \psi(e_k) - \ln f_k, \quad (46)$$

$$E[\ln \pi_t^{v'}] = \psi(r_t^v) - \psi(r_t^v + w_t^v), \quad (47)$$

$$E[\ln(1 - \pi_t^{v'})] = \psi(w_t^v) - \psi(r_t^v + w_t^v), \quad (48)$$

$$E[\ln \pi_t^v] = E[\ln \pi_t^{v'}] + \sum_{s=1}^{t-1} E[\ln(1 - \pi_s^{v'})], \quad (49)$$

$$E[\ln \beta_k'] = \psi(s_k^1) - \psi(s_k^1 + s_k^2), \quad (50)$$

$$E[\ln(1 - \beta_k')] = \psi(s_k^2) - \psi(s_k^1 + s_k^2), \quad (51)$$

$$E[\ln \beta_k] = E[\ln \beta_k'] + \sum_{l=1}^{k-1} E[\ln(1 - \beta_l')], \quad (52)$$

$$E[\mathbf{L}_{i \cdot}^v \mathbf{L}_{i \cdot}^{v \cdot T}] = \boldsymbol{\mu}_i^v \boldsymbol{\mu}_i^{v \cdot T} + \boldsymbol{\Sigma}_i^v, \quad (53)$$

$$E[\mathbf{R}_{\cdot j} \mathbf{R}_{\cdot j}^T] = \boldsymbol{\mu}_j \boldsymbol{\mu}_j^T + \boldsymbol{\Sigma}_j, \quad (54)$$

$$E[\mathbf{L}_{r \cdot}^v \mathbf{L}_{r \cdot}^{v \cdot T}] = \sum_i (\boldsymbol{\mu}_{ir}^v)^2 + (\boldsymbol{\Sigma}_r^v)_{rr}, \quad (55)$$

$$E[\mathbf{R}_{r \cdot}^T \mathbf{R}_{r \cdot}] = \sum_j (\boldsymbol{\mu}_{jr})^2 + (\boldsymbol{\Sigma}_r)_{rr}, \quad (56)$$

$$\begin{aligned} E[(x_{ij}^v - \mathbf{L}_{i \cdot}^v \mathbf{R}_{j \cdot}^v)^2] &= x_{ij}^{v \cdot 2} - 2x_{ij}^v \boldsymbol{\mu}_i^v \boldsymbol{\mu}_j^T + \boldsymbol{\Sigma}_r^v \\ &+ \text{tr} \left(E[\mathbf{L}_{i \cdot}^v \mathbf{L}_{i \cdot}^{v \cdot T}] E[\mathbf{R}_{\cdot j} \mathbf{R}_{\cdot j}^T] \right), \end{aligned} \quad (57)$$

where μ_{ir}^v and μ_{jr} represent the r th element of vectors μ_i^v and μ_j , respectively, and $\psi(\cdot)$ is the Digamma function defined by $\psi(x) = \frac{d}{dx} \ln \Gamma(x)$

II. VALIDATION EXPERIMENTS ON NON-I.I.D. ASSUMPTION

A. Two Baseline Methods

In the main text, we assume that the noise of the practical multi-view data have three characteristics, i.e., complex, non-identical and non-independent. Previous noise modeling literatures [1]–[3] have comprehensively validated the effectiveness of the complex noise assumption in different real applications. We thus want to further test the necessity of the non-identical and non-independent assumptions in our noise modeling. In order to demonstrate the benefit of these two assumptions in model performance, we design two noise models as the baseline methods to be compared with the NIID-MSL model. The first one fits the noise using one single DPGMM for all the views (complex but i.i.d.) and the second one assumes different DPGMMs for all views of data (complex, non-identical, but independent). Since the latent subspace modeling part of these two baselines are the same with the NIID-MSL as shown in Eq. (3), we only list the noise modeling parts of both baselines as follows:

Baseline 1:

$$\xi_k \sim \text{Gam}(e_0, f_0), \quad e_{ij}^v \sim \mathcal{N}(0, (\xi_{z_{ij}}^v)^{-1}), \quad (58a)$$

$$z_{ij}^v \sim \text{Multi}(\boldsymbol{\pi}), \quad \pi_k = \pi_k' \prod_l^{k-1} (1 - \pi_l'), \quad (58b)$$

$$\pi_k' \sim \text{Beta}(1, \gamma), \quad \gamma \sim \text{Gam}(m_0, n_0). \quad (58c)$$

Baseline 2:

$$\xi_k \sim \text{Gam}(e_0, f_0), \quad e_{ij}^v \sim \mathcal{N}(0, (\xi_{z_{ij}}^v)^{-1}), \quad (59a)$$

$$z_{ij}^v \sim \text{Multi}(\boldsymbol{\pi}^v), \quad \pi_k^v = \pi_k'^v \prod_l^{k-1} (1 - \pi_l'^v), \quad (59b)$$

$$\pi_k'^v \sim \text{Beta}(1, \gamma^v), \quad \gamma^v \sim \text{Gam}(m_0, n_0). \quad (59c)$$

It should be noted that the proposed NIID-MSL method can be easily transformed into both of its degenerated versions, including the i.i.d. noise version by using a single DPGMM to fit the noise of all the views, and non-identical but independent noise version by assuming a different DPGMM for each view.

B. Experimental Results

We compare our proposed NIID-MSL methods with two baseline methods to validate the rationality and necessity of our non-identical and non-independent assumptions on noises of the multi-view data. Experiments were carried out on the ‘No noise’ case in the face image recovery experiments and the ‘Wallflower’ datasets in the foreground detection experiments from the RGB videos, respectively. Both represent the typical cases of real noises in practical multi-view data.

Table I and Table II list the quantitative comparison of the face image recovery experiments and foreground detection experiments, in terms of RRSE/RRAE and F-Measure,

TABLE I
QUANTITATIVE PERFORMANCE COMPARISON OF NIID-MSL AND TWO BASELINE METHODS ON THE CMU MULTI-PIE FACE DATASETS WITHOUT SYNTHETIC NOISE. IN EACH SERIES OF EXPERIMENTS, THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Index	Methods		
	Baseline 1	Baseline 2	NIID-MSL
RRSE	0.012	0.011	0.009
RRAE	0.068	0.068	0.067

TABLE II
QUANTITATIVE PERFORMANCE COMPARISON OF NIID-MSL AND TWO BASELINE METHODS ON THE WALLFLOWER DATASET. IN EACH SERIES OF EXPERIMENTS, THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Video	Methods		
	Baseline 1	Baseline 2	NIID-MSL
Bootstrapping	0.73	0.73	0.73
Camouflage	0.72	0.72	0.74
Apertu	0.96	0.96	0.96
SwitchLight	0.68	0.68	0.69
TimeOfDay	0.76	0.77	0.76
WavingTrees	0.72	0.66	0.91
Mean	0.76	0.76	0.80

respectively. From the results listed in the tables, we can easily observe the superiority of the proposed method. This supports the necessity of such non-identical and non-independent considerations on the noise modeling of real multi-view data.

III. SENSITIVITY TEST OF THE RANK PARAMETER IN FACE EXPERIMENTS

In Section VI.B of the main text, we conduct some experiments on face image recovery, in which the rank is set as 15. We try to give more explanations on the physical meanings under this parameter setting in this section.

After concatenating all the V views of the data $\mathbf{X} = \{\mathbf{X}^v\}_{v=1}^V$ along the vertical direction, Eq. (11) in the main text, i.e.,

$$\mathbf{X}^v = \mathbf{L}^v \mathbf{R} = \sum_{r=1}^l \mathbf{L}_r^v \mathbf{R}_{r.r.} \quad (60)$$

This is equivalent to

$$\mathbf{X} = \mathbf{L}\mathbf{R} = \sum_{r=1}^l \mathbf{L}_{r.r.} \mathbf{R}_{r.r.}, \quad (61)$$

where $\mathbf{X} = \begin{bmatrix} \mathbf{X}^1 \\ \mathbf{X}^2 \\ \vdots \\ \mathbf{X}^V \end{bmatrix}$, $\mathbf{L} = \begin{bmatrix} \mathbf{L}^1 \\ \mathbf{L}^2 \\ \vdots \\ \mathbf{L}^V \end{bmatrix}$. The above Eq. (61) indicates that the rank corresponds to the number of bases $\{\mathbf{L}_{r.r.}\}_{r=1}^l$ for generating the observed data with coefficients \mathbf{R} .

In the face image recovery experiments, we employed the CMU Multi-PIE face dataset, which includes 337 subjects with multiple poses and expressions. Due to memory limitation of our computer, 100 subjects were randomly selected and each subject contains one face image at 5 different angles with size 128×96 . Under the notation assumption of our main text, $d = 12288 = 128 * 96$, $n = 100$ and $V = 5$. According to Eq. (60) or (61), it is easy to see that l basis

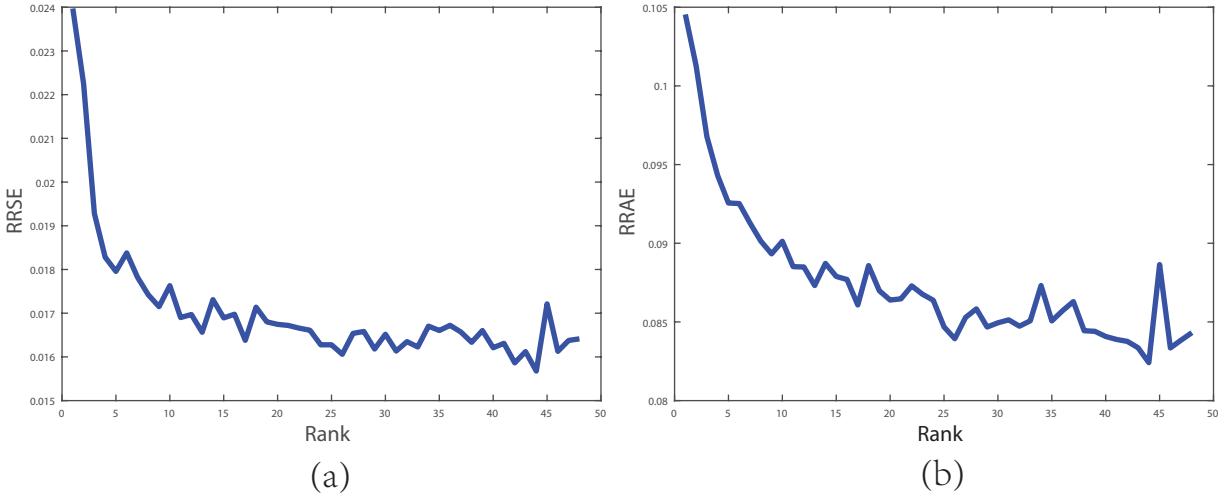


Fig. 1. The curves of performance change tendency of the NIID-MSL method under different rank values in the CMU Multi-PIE face image recovery experiment in terms of (a) RRSE (Relative Reconstruction Square Error) and (b) RRAE (Relative Reconstruction Absolute Error).

face images in \mathbf{L} are used to construct all of the face images through linear combination. Intuitively, if the rank l is too small, the recovered face images will be too similar and the specific characteristic of each face tends to be insufficiently delivered. On the contrary, if we set this parameter too large, the basis matrix \mathbf{L} will be redundant and it inclines to absorb unexpected noises and hamper the final performance. Thus the parameter should be specified as a moderate value to guarantee a satisfactory performance of our method. Actually, we have tried the parameter within a wide range (from 3 to 50), and observed that our method can perform consistently well from around 10 to 40. We thus easily set it as 15 for all our face experiments, and all subspace-learning-based methods, including ours, can attain good performance both quantitatively (in terms of RRSE and RRAE) and visually. To better clarify this point, we show the changing tendency in performance of our method in the CMU Multi-PIE face image recovery experiments with the rank l varying from 3 to 50 in terms of RRSE and RRAE, respectively, in Fig. 1(a) and (b). It can be easily observed that both tendency curves are relatively stable within the range from 10 to 40.

IV. AUC PERFORMANCE DEMONSTRATION IN FOREGROUND DETECTION EXPERIMENTS ON RGB DATA

In this section we show the AUC curves of all competing methods implemented on the WallFlower and I2R datasets in Fig. 2 and Fig. 3, respectively, to more comprehensively evaluate the effectiveness of our proposed method. The relatively better performance of the proposed method can be easily observed from these figures, which complies with the F-measure results listed in Table IV and Table V of the manuscript.

REFERENCES

- [1] D. Meng and F. De La Torre, “Robust matrix factorization with unknown noise,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1337–1344.
- [2] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang, “Robust principal component analysis with complex noise,” in *International conference on machine learning*, 2014, pp. 55–63.
- [3] X. Cao, Y. Chen, Q. Zhao, D. Meng, Y. Wang, D. Wang, and Z. Xu, “Low-rank matrix factorization under general mixture noise distributions,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1493–1501.

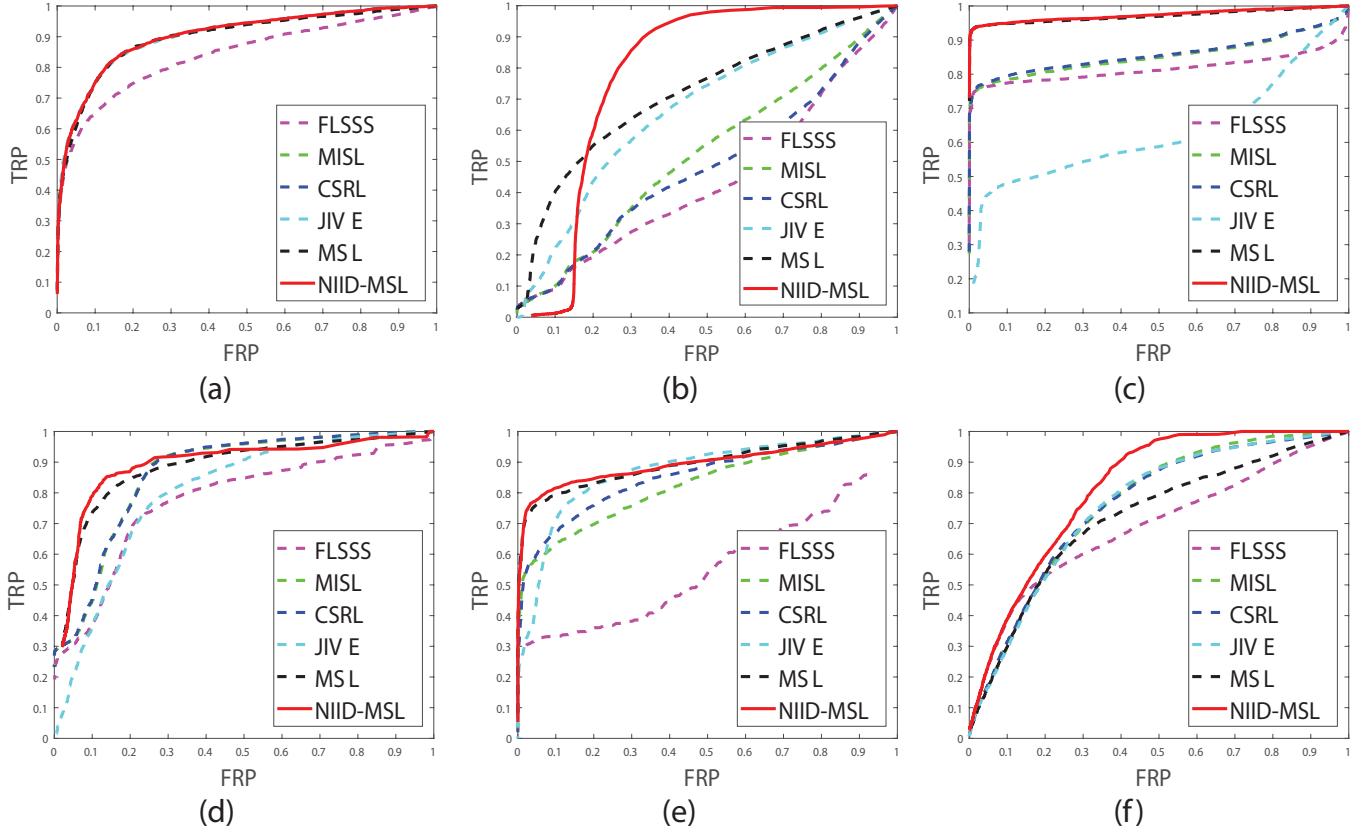


Fig. 2. The AUC curves of different methods on the WallFlower dataset: (a) Bootstrapping, (b) Camouflage, (c) Apertu, (d) SwitchLight, (e) TimeOfDay, (f) WavingTrees.

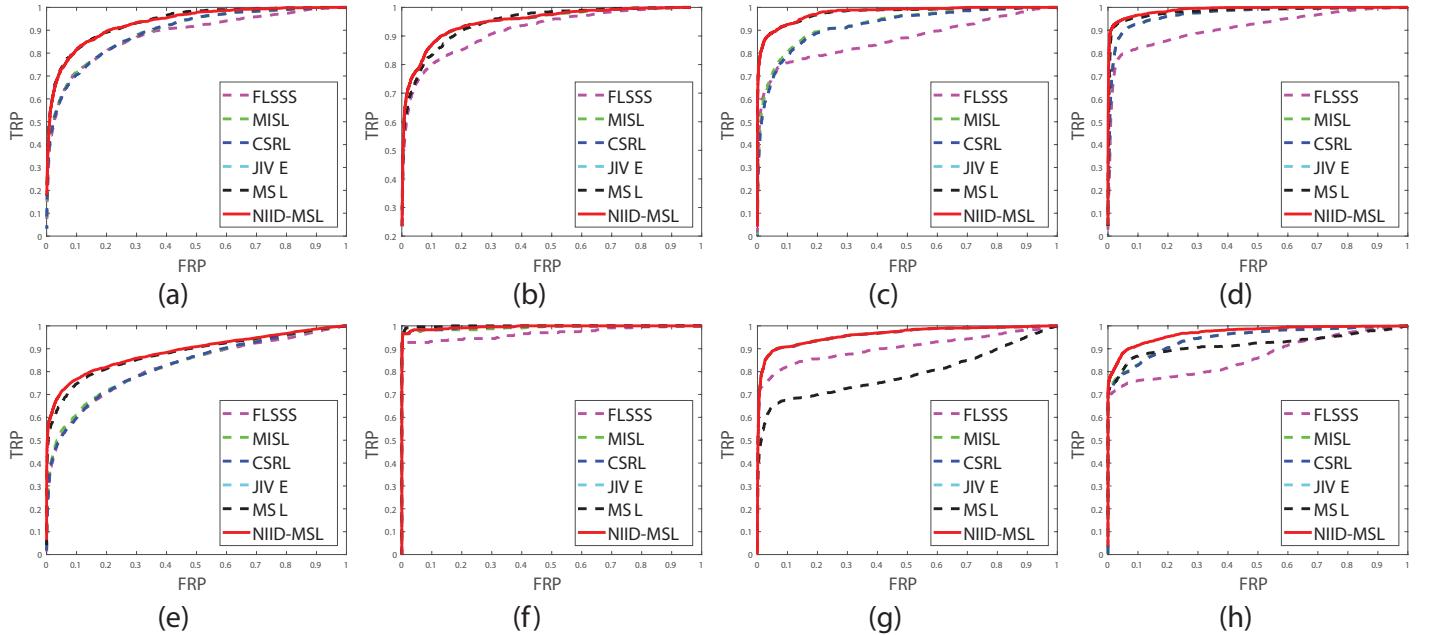


Fig. 3. The AUC curves of different methods on the I2R dataset: (a) Campus, (b) Escalator, (c) Fountain, (d) Curtain, (e) Hall, (f) Lobby, (g) ShoppingMall, (h) WaterSurface.